

Joseph R. Saveri (State Bar No. 130064)  
Cadio Zirpoli (State Bar No. 179108)  
Christopher K.L. Young (State Bar No. 318371)  
Holden Benon (State Bar No. 325847)  
Aaron Cera (State Bar No. 351163)  
Kathleen J. McMahon (State Bar No. 340007)

Margaux Poueymirou (State Bar No. 356000)  
**JOSEPH SAVERI LAW FIRM, LLP**  
601 California Street, Suite 105005  
San Francisco, California 94108  
Telephone: (415) 500-6800  
Facsimile: (415) 395-9940  
Email: jsaveri@saverilawfirm.com  
czirpoli@saverilawfirm.com  
cyoung@saverilawfirm.com  
hbenon@saverilawfirm.com  
acera@saverilawfirm.com  
kmcmahon@saverilawfirm.com  
mpoueymirou@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)  
1920 Hillhurst Avenue, #406  
Los Angeles, CA 90027  
Telephone: (323) 968-2632  
Facsimile: (415) 395-9940  
Email: mb@buttericklaw.com

*Counsel for Individual and Representative  
Plaintiffs and the Proposed Class*

[Additional Counsel Listed on Signature Page]

Matthew Butterick (State Bar No. 250953)  
1920 Hillhurst Avenue, #406  
Los Angeles, CA 90027  
Telephone: (323) 968-2632  
Facsimile: (415) 395-9940  
Email: mb@buttericklaw.com

Bryan L. Clobes (*pro hac vice*)  
Alexander J. Sweatman (pro hac vice anticipated)  
**CAFFERTY CLOBES MERIWETHER  
& SPRENGEL LLP**  
135 South LaSalle Street, Suite 3210  
Chicago, IL 60603  
205 N. Monroe Street  
Media, PA 19063  
Telephone: (215) 864-2800  
Email: bclobes@caffertyclobes.com  
asweatman@caffertyclobes.com

David A. Straite (pro hac vice)  
**DICELLO LEVITT LLP**  
475 Lexington Avenue, Suite 1001  
New York, NY 10017  
Telephone: (646) 933-1000  
Email: dstraite@dicellolevitt.com

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
SAN FRANCISCO DIVISION**

Richard Kadrey, Sarah Silverman, Christopher  
Golden, ~~Michael Chabon~~, Ta-Nehisi Coates, Junot  
Díaz, Andrew Sean Greer, David Henry Hwang,  
Matthew Klam, Laura Lippman, Rachel Louise  
Snyder, ~~Ayelet Waldman~~, and Jacqueline Woodson,  
and Lysa TerKeurst,

*Individual and Representative Plaintiffs,*

v.

Meta Platforms, Inc., a Delaware corporation;

Case No. 3:23-cv-03417-VC

**~~FIRST~~SECOND CONSOLIDATED  
AMENDED COMPLAINT**

**Class Action**

**Demand for Jury Trial**

*Defendant.*

Plaintiffs Richard Kadrey, Sarah Silverman, Christopher Golden, ~~Michael Chabon~~, Ta-Nehisi Coates, Junot Díaz, Andrew Sean Greer, David Henry Hwang, Matthew Klam, Laura Lippman, Rachel Louise Snyder, ~~Ayelet Waldman~~, and Jacqueline Woodson and Lysa TerKeurst (“Plaintiffs”), on behalf of themselves and all others similarly situated, bring this Class Action Complaint (the “Complaint”) against Defendant Meta Platforms, Inc. (“Meta”).

## I. OVERVIEW

1. “Llama” is the name given to a series of large language models created and maintained by Meta. A *large language model* is an AI software program designed to emit convincingly naturalistic text outputs in response to user prompts.

2. Rather than being programmed in the traditional way, a large language model is “trained” by copying massive amounts of text and extracting expressive information from it. This body of text is called the *training dataset*.

3. A large language model’s output is therefore entirely and uniquely reliant on the material in its training dataset. Every time it assembles a text output, the model relies on the information it extracted from its training dataset. Thus, the decisions about what textual information to include in the training dataset are deliberate and important choices.

4. Plaintiffs and Class members are authors of books. Plaintiffs and Class members have copyrights in the books they published. Plaintiffs and Class members did not consent to the use of their copyrighted books as training material for Llama 1 and Llama 2.

5. Nonetheless, their copyrighted materials were copied and ingested as part of training Llama 1 and Llama 2. Many of Plaintiffs’ copyrighted books appear in the dataset that Meta has admitted to using to train Llama 1. On information and belief, Plaintiffs’ copyrighted books also appear in the dataset that Meta used to train Llama 2.

## II. JURISDICTION AND VENUE

6. This Court has subject matter jurisdiction under 28 U.S.C. § 1331 because this case arises under the Copyright Act (17 U.S.C. § 101 *et seq.*).

7. Jurisdiction and venue is proper in this judicial district under 28 U.S.C. § 1391(c)(2) because Defendant Meta Platforms, Inc. (“Meta”) is headquartered in this district, and thus a substantial part of the events giving rise to the claim occurred in this district; and because a substantial part of the events giving rise to Plaintiffs’ claims occurred in this District, and a substantial portion of the affected interstate trade and commerce was carried out in this District. Defendant has transacted business, maintained substantial contacts, and/or committed overt acts in furtherance of the illegal scheme and conspiracy throughout the United States, including in this District. Defendant’s conduct has had the intended and foreseeable effect of causing injury to persons residing in, located in, or doing business throughout the United States, including in this District.

8. Under Civil Local Rule 3.2(d), assignment of this case to the San Francisco Division is proper because Meta is headquartered in San Mateo County, where a substantial part of the events giving rise to the claim occurred, a substantial amount part of the events giving rise to Plaintiffs’ claims and the interstate trade and commerce involved and affected by Defendant’s conduct giving rise to the claims herein occurred in this Division.

## III. PARTIES

### A. Plaintiffs

9. Plaintiff Richard Kadrey is a writer who lives in Pennsylvania and owns registered copyrights in multiple works, including *Sandman Slim*.

10. Plaintiff Sarah Silverman is a writer and performer who lives in California and owns a registered copyright in one work, called *The Bedwetter*.

11. Plaintiff Christopher Golden is a writer who lives in Massachusetts and owns registered copyrights in multiple works, including *Ararat*.

~~12. Plaintiff Michael Chabon is an author who lives in California and owns registered copyrights in multiple works, including *The Amazing Adventures of Kavalier & Clay*.~~

~~13.~~12. Plaintiff Ta-Nehisi Coates is an author who lives in New York and owns registered copyrights in multiple works, including *The Beautiful Struggle*.

~~14.~~13. Plaintiff Junot Díaz is an author who lives in Massachusetts and owns registered copyrights in multiple works, including *Drown*.

~~15.~~14. Plaintiff Andrew Sean Greer is an author who lives in California and owns registered copyrights in multiple works, including *The Confessions of Max Tivoli*.

~~16.~~15. Plaintiff David Henry Hwang is a playwright and screenwriter who lives in New York and owns registered copyrights in multiple works, including *M. Butterfly*.

~~17.~~16. Plaintiff Matthew Klam is an author who lives in Washington, D.C. and owns registered copyrights in multiple works, including *Who is Rich?*

~~18.~~17. Plaintiff Laura Lippman is an author who lives in Maryland and owns registered copyrights in multiple works, including *After I'm Gone*.

~~19.~~18. Plaintiff Rachel Louise Snyder is an author who lives in Washington, D.C. and owns registered copyrights in multiple works, including *No Visible Bruises: What We Don't Know About Domestic Violence Can Kill Us*.

~~20. Plaintiff Ayelet Waldman is an author and screen and television writer who lives in California and owns registered copyrights in multiple works, including *Love and Other Impossible Pursuits*.~~

19. Plaintiff Jacqueline Woodson is an author who lives in New York and owns registered copyrights in multiple works, including *Brown Girl Dreaming*.

~~21.20.~~ Plaintiff ~~Ayelet Waldman~~ Lysa TerKeurst is an author ~~and screen and television writer~~ who lives in ~~California~~ St. Johns County, Florida, and owns registered copyrights in multiple works, including ~~Love and Other Impossible Pursuits. Uninvited.~~

~~22.21.~~ A nonexhaustive list of registered copyrights owned by Plaintiffs is included as Exhibit A.

**B. Defendant**

~~23.22.~~ Defendant Meta is a Delaware corporation with its principal place of business at 1601 Willow Road, Menlo Park, California 94025.

**IV. AGENTS AND CO-CONSPIRATORS**

~~24.23.~~ The unlawful acts alleged against the Defendant in this class action complaint were authorized, ordered, or performed by the Defendant's respective officers, agents, employees, representatives, or shareholders while actively engaged in the management, direction, or control of the Defendant's businesses or affairs. The Defendant's agents operated under the explicit and apparent authority of their principals. Each Defendant, and its subsidiaries, affiliates, and agents operated as a single unified entity.

~~25.24.~~ Various persons and/or firms not named as Defendants may have participated as co-conspirators in the violations alleged herein and may have performed acts and made statements in furtherance thereof. Each acted as the principal, agent, or joint venture of, or for other Defendants with respect to the acts, violations, and common course of conduct alleged herein.

**V. FACTUAL ALLEGATIONS**

~~26.25.~~ Meta is a diversified internet company that creates, markets, and sells software and hardware technology products, including Facebook, Instagram, and Horizon Worlds. Meta also has a large artificial-intelligence group called Meta AI that creates and distributes artificial-intelligence software products.

~~27.26.~~ *Artificial intelligence* is commonly abbreviated “AI.” AI software is designed to algorithmically simulate human reasoning or inference, often using statistical methods.

~~28.27.~~ In February 2023, Meta released the initial version of an AI product called LLaMA, though Meta has since revised its spelling to “Llama.” Llama is a series of *large language models*. A large language model (or “LLM” for short) is AI software designed to parse and emit natural language. Though a large language model is a software program, it is not created the way most software programs are—that is, by human software engineers writing code. Rather, a large language model is “trained” by copying massive amounts of text from various sources and feeding these copies into the model. This corpus of input material is called the *training dataset*. During training, the large language model copies each piece of text in the training dataset and extracts expressive information from it. The large language model progressively adjusts its output to more closely resemble the sequences of words copied from the training dataset. Once the large language model has copied and ingested all this text, it is able to emit convincing simulations of natural written language as it appears in the training dataset.

~~29.28.~~ Much of the material in Meta’s training dataset, however, comes from copyrighted works—including books written by Plaintiffs—that were copied by Meta without consent, without credit, and without compensation.

~~30.29.~~ The first version of Llama, called Llama 1, was trained between December 2022 and February 2023.

~~31.30.~~ In February 2023, Meta introduced Llama 1 in a paper called “[Llama 1]: Open and Efficient Foundation Language Models”<sup>1</sup> (the “Llama 1 Paper”). In the Llama 1 Paper, Meta describes the Llama 1 training dataset as “a large quantity of textual data” that was chosen because it was “publicly available, and compatible with open sourcing.”

---

<sup>1</sup> <https://arxiv.org/pdf/2302.13971.pdf>

~~32.31.~~ 33.31. *Open sourcing* refers to putting data under a permissive style of copyright license called an *open-source license*. Copyrighted materials, however, are not ordinarily “compatible with open sourcing” unless and until the copyright owner first places the material under an open-source license, thereby enabling others to do so later.

~~33.32.~~ 34.32. In the Llama 1 Paper, in a table describing the composition of the Llama 1 training dataset, Meta notes that 85 gigabytes of the training data comes from a category called “Books.” Meta further elaborates that “Books” comprises the text of books from two internet sources: (1) Project Gutenberg, an online archive of approximately 70,000 books that are out of copyright, and (2) “the Books3 section of ThePile . . . a publicly available dataset for training large language models.” The Llama 1 Paper does not further describe the contents of Books3 or The Pile.

~~34.33.~~ 35.33. But that information is available elsewhere. The Pile is a dataset assembled by a research organization called EleutherAI. In December 2020, EleutherAI discussed this dataset in a paper called “The Pile: An 800GB Dataset of Diverse Text for Language Modeling”<sup>2</sup> (The “EleutherAI Paper”).

~~35.34.~~ 36.34. The EleutherAI Paper reveals that the Books3 dataset comprises 108 gigabytes of data, or approximately 12% of the dataset, making it the third largest component of The Pile by size.

~~36.35.~~ 37.35. The EleutherAI Paper describes the contents of Books3:

Books3 is a dataset of books derived from a copy of the contents of the Bibliotik private tracker . . . Bibliotik consists of a mix of fiction and nonfiction books and is almost an order of magnitude larger than our next largest book dataset (BookCorpus2). We included Bibliotik because books are invaluable for long-range context modeling research and coherent storytelling.

~~37.36.~~ 38.36. Bibliotik is one of a number of notorious “shadow library” websites that also includes Library Genesis (aka LibGen), Z-Library (aka B-ok), Sci-Hub, and Anna’s Archive. The books and

---

<sup>2</sup> <https://arxiv.org/pdf/2101.00027.pdf>



other materials aggregated by these websites have also been available in bulk via torrent systems. These shadow libraries have long been of interest to the AI-training community because of the large quantity of copyrighted material they host. For that reason, these shadow libraries are also flagrantly illegal.

~~38.37.~~ The person who assembled the Books3 dataset, Shawn Presser, has confirmed in public statements that it represents “all of Bibliotik”<sup>3</sup> and contains 196,640 books.

~~39.38.~~ Many of Plaintiffs’ books appear in the Books3 dataset. A list of Plaintiffs’ books currently known to exist in the Books3 dataset is attached as Exhibit B. Together, these books are referred to as the **Infringed Works**.

~~40.39.~~ Shawn Presser posts comments on the website Hacker News (<https://news.ycombinator.com>) under the name “sillysaurusx.” In response to another commenter asking about the origin of Books3 in 2020, Presser said, “It was bibliotik. . . . The llama folks [meaning Meta] had to [remove duplicates from] the books themselves. . . . Basically, the-eye.eu was at one point hosting all of bibliotik, so I downloaded all the epubs [the native format for electronic books] and converted them to text. I still have those epubs (incidentally thanks to Carmack, who through a convoluted process managed to save the them and send them to me via snail mail).”<sup>4</sup> On information and belief, the only “Carmack” who could be identified mononymously on Hacker News is prominent software engineer John Carmack, who was employed by Meta during the events described by Presser.

~~41.40.~~ Until August 2023, EleutherAI facilitated the download of copies of Books3 through its website (<https://pile.eleuther.ai/>) by linking to a second site called The Eye (the-eye.eu). In August

<sup>3</sup> <https://twitter.com/theshawwn/status/1320282149329784833>

<sup>4</sup> <https://news.ycombinator.com/item?id=36197731>

2023, the Books3 dataset was removed from The Eye in response to a takedown notice by the Danish Rights Alliance.

42.41. Until October 2023, the Books3 dataset was also available from a popular AI project hosting service called Hugging Face ([https://huggingface.co/datasets/the\\_pile\\_books3](https://huggingface.co/datasets/the_pile_books3)). In October 2023, the Books3 dataset was removed from Hugging Face, with a message that it “is defunct and no longer accessible due to reported copyright infringement.”

43.42. On information and belief, the Books3 dataset is still circulating on the public internet and can be downloaded by those sufficiently motivated.

44.43. In the Llama 1 Paper, Meta says it copied material for the Llama 1 training dataset that was “publicly available.” Importantly, however, “publicly available” does not mean “public domain.” A work in the public domain is not protected by copyright. A work that is publicly available, on the other hand, may still be protected by copyright and other intellectual-property laws.

45.44. Meta is well aware of this distinction. In the Llama 1 Paper, Meta describes Project Gutenberg as comprising “books that are in the *public domain*,” whereas it acknowledges Books3 was merely “a *publicly available* dataset” (emphases added).

46.45. Still, even “publicly available” is a misleading description of Books3. The books in Books3 were not put there by the copyright owners, including Plaintiffs. Rather, their books became “publicly available” via Books3 only because of the willful efforts of John Carmack, Shawn Presser, and EleutherAI to copy and distribute them for free without the authorization of Plaintiffs.

47.46. Before Meta used Books3 for training its language models, it had already publicly acknowledged the legal problems with Books3.

48.47. In November 2020, Meta AI researcher Tim Dettmers initiated a conversation on the EleutherAI public Discord server about Meta’s interest in using The Pile as training data.

~~49-48.~~ Dettmers said, “It is really great work that you are doing with The Pile. Having a large public dataset that is easily accessible was long overdue!” A couple minutes later, Dettmers added, “A colleague of mine and I wanted to use [The Pile] dataset in our research and wondering what would be the best way to build on your research efforts.”

~~50-49.~~ Leo Gao, an AI researcher affiliated with EleutherAI, exchanged a number of messages with Dettmers, and then said, “any downstream use of the data can begin right away.”

~~51-50.~~ Dettmers asked, “in other words you would be happy if we are already using the data?”

~~52-51.~~ Gao responded, “yup, you can start now.”

~~53-52.~~ Dettmers asked, “It is okay to download [The Pile] from the server that you linked?”

~~54-53.~~ Gao responded, “yeah, go right ahead.”

~~55-54.~~ Dettmers then asked, “[REDACTED]”. Do you have a sense if there would be any legal concerns with parts of the data?”

~~56-55.~~ Gao responded, “[I] believe that merely training on it should fall under fair use ... Stella Biderman [another AI researcher affiliated with EleutherAI] might have a bit more to say [with respect to] legality.”

~~57-56.~~ Dettmers said, “Sounds good! [REDACTED]”  
[REDACTED]  
[REDACTED]  
[REDACTED]  
[REDACTED]  
[REDACTED]

~~58-57.~~ Responding to Dettmers, Gao said “[I]’m so glad that you like [The Pile] and are interested in using it in your work and [I]’d definitely be down to talk about any potential legal problems in the future.”

~~59-58.~~ Responding to Dettmers, Stella Biderman said, “Happy to chat about legal questions you have. tl;dr your legal dept is most likely to be worried about books3 which contains the text of books with active copyrights. . . . In the US this is all a legal grey area because of a lack of court rulings, but there’s a very strong case for free use, even with books3.”

~~60-59.~~ In December 2020, Dettmers posted on the EleutherAI Discord server: [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

~~61-60.~~ In January 2021, Dettmers posted on the EleutherAI Discord server: “At Facebook [since renamed Meta] there are a lot of people interested in working with [T]he [P]ile, including myself, [REDACTED] Would there be interest in working on this together, [REDACTED]

[REDACTED]

~~62-61.~~ Later that day, Dettmers added, [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

~~63-62.~~ [REDACTED]

[REDACTED] Nevertheless, between December 2022 and February 2023, Meta still included Books3 in the training dataset for Llama 1, causing the books in Books3 to be copied and ingested during the training process, and published Llama 1.

~~64-63.~~ In August 2023, EleutherAI removed from public view these conversations between Tim Dettmers, Stella Biderman, and Leo Gao concerning The Pile and Books3.

~~65-64.~~ At the launch of the Llama 1 language models in February 2023, Meta made those models selectively available to organizations that requested access, saying:

To maintain integrity and prevent misuse, we are releasing our model under a noncommercial license focused on research use cases. Access to the model will be granted on a case-by-case basis to academic researchers; those affiliated with organizations in government, civil society, and academia; and industry research laboratories around the world. People interested in applying for access can find the link to the application in our research paper.

~~66-65.~~ Meta has not disclosed what criteria it used to decide who was eligible to receive the Llama 1 language models, nor who actually received them, nor whether Meta in fact adhered to its stated criteria. On information and belief, Meta has in fact distributed the Llama 1 models to certain people and entities, continues to do so, and has benefited financially from these acts. Meta would later say that it “received unprecedented interest in the Llama 1 model we released for the research community—more than 100,000 individuals and organizations have applied for access to Llama 1 and tens of thousands are now using it to innovate.” This implies that Meta’s original suggestion that Llama 1 was focused on research was pretextual. Rather, Llama 1 was always intended to either become commercially available and lucrative, or as a precursor to another product that would be commercially available and lucrative.

~~67-66.~~ In March 2023, the Llama 1 language models were leaked to a public internet site and have continued to circulate. Meta has not disclosed what role it had, if any, in the leak.

~~68-67.~~ Later in March 2023, Meta issued a DMCA takedown notice to a programmer on GitHub who had released a tool that helped users download the leaked Llama 1 language models. In the notice, Meta asserted copyright over the Llama 1 language models.

~~69-68.~~ Between January and July 2023, Meta trained the successor to the Llama 1 language models, called Llama 2. On information and belief, Llama 2 was also trained on Books3, because the training period for Llama 2 (January–July 2023) overlapped with the training for Llama 1 (December 2022–February 2023), and took place before the initial complaint in this action was filed.

~~70-69.~~ Meta released the Llama 2 models in July 2023, after the initial complaint in this action was filed. Information about Llama 2 is available in a research paper released on July 19, 2023 called “Llama 2: Open Foundation and Fine-Tuned Chat Models”<sup>5</sup> (the “Llama 2 Paper”).

~~71-70.~~ With Llama 2, Meta abandoned all pretext of non-commercial purpose. According to the FAQ for Llama 2,<sup>6</sup> “Llama 2 is . . . available under a permissive commercial license, whereas Llama 1 was limited to non-commercial use.”

~~72-71.~~ In contrast to Llama 1, Meta chose not to reveal the training datasets for Llama 2. According to the FAQ for Llama 2, “[Q:] Where did the data come from to train the models? . . . [A:] A combination of sources are used for training. These sources include information that is publicly available online and annotated data to train our models. . . . [Q:] Why are you not sharing the training datasets for Llama 2? . . . [A:] data mixes are intentionally withheld for competitive reasons.”

~~73-72.~~ This explanation, however, is likely pretextual. As explained in the Llama 2 Paper, Llama 2—like Llama 1—was also trained on a “mix of publicly available data.” A more plausible

---

<sup>5</sup> <https://arxiv.org/pdf/2307.09288.pdf>

<sup>6</sup> <https://ai.meta.com/llama/faq/>

explanation for Meta’s decision to conceal its training data is to avoid scrutiny by those whose copyrighted works were copied and ingested during the training process for Llama 2.

~~74.73.~~ On information and belief, a key reason Meta chose not to share the training dataset for Llama 2 was to avoid litigation from using copyrighted materials for training that Meta had previously determined to be legally problematic. Indeed, as Meta acknowledged in its Form 10-Q filing on October 26, 2023, Meta is the subject of lawsuits which are challenging the “alleged use of copyright-protected content to train our [Meta’s] AI models” and “any negative outcome from any such lawsuits could result in payments of substantial monetary damages or fines, or undesirable changes to our products or business practices, and accordingly our business, financial condition, or results of operations could be materially and adversely affected.”

~~75.74.~~ The Llama 2 Paper warns that the “Llama 2 models should be used carefully and deployed only after significant safety tuning is applied” because “Llama 2 does not outperform other models on toxicity metrics.” The term *toxicity metrics* refers to measurements of a language model’s propensity to emit output that is offensive, dangerous, or harmful. Meta “speculate[s] that [Llama’s comparatively poor performance on toxicity metrics] may be because we refrained from aggressively filtering the [training] data. ... We reiterate that this ... choice does imply that additional safety mitigations should be applied before deployment.”

~~76.75.~~ Despite these severe warnings, Meta still made the Llama 2 models available for free, to anyone, for commercial or noncommercial purposes.

## VI. CLAIM FOR RELIEF

### Direct Copyright Infringement 17 U.S.C. § 101 et seq.

~~77.76.~~ Plaintiffs incorporate by reference the preceding factual allegations.

~~78.77.~~ As the owners of the registered copyrights in the Infringed Works, Plaintiffs hold the exclusive rights to those books under 17 U.S.C. § 106.

~~79.78.~~ To train the Llama 1 and Llama 2 language models, Meta copied the Books3 dataset, which includes the Infringed Works.

~~80.79.~~ Plaintiffs never authorized Meta to make copies of their Infringed Works, make derivative works, publicly display copies (or derivative works), or distribute copies (or derivative works). All those rights belong exclusively to Plaintiffs under copyright law.

~~81.80.~~ Meta made copies of the Infringed Works during the training process of the Llama 1 and Llama 2 language models without Plaintiffs' permission.

~~82.81.~~ Plaintiffs have been injured by Meta's acts of direct copyright infringement. Plaintiffs are entitled to statutory damages, actual damages, restitution of profits, and other remedies provided law.

## VII. CLASS ALLEGATIONS

### A. Class Definition

~~83.82.~~ Plaintiffs bring this action for damages and injunctive relief as a class action under Federal Rules of Civil Procedure 23(a), 23(b)(2), and 23(b)(3), on behalf of the following Class:

**All persons or entities domiciled in the United States that own a United States copyright in any work that was used as training data for any version of the Llama language models between July 7, 2020 and the present (the "Class Period").**

~~84.83.~~ This Class definition excludes:

- a. Defendant named herein;
- b. any of the Defendant's co-conspirators;
- c. any of Defendant's parent companies, subsidiaries, and affiliates;
- d. any of Defendant's officers, directors, management, employees, subsidiaries, affiliates, or agents;
- e. all governmental entities; and



- f. the judges and chambers staff in this case, as well as any members of their immediate families.

**B. Numerosity**

~~85~~84. Plaintiffs do not know the exact number of members in the Class. This information is in the exclusive control of Defendant. On information and belief, there are at least thousands of members in the Class geographically dispersed throughout the United States. Therefore, joinder of all members of the Class in the prosecution of this action is impracticable.

**C. Typicality**

~~86~~85. Plaintiffs' claims are typical of the claims of other members of the Class because Plaintiffs and all members of the Class were damaged by the same wrongful conduct of Defendant as alleged herein, and the relief sought herein is common to all members of the Class.

**D. Adequacy**

~~87~~86. Plaintiffs will fairly and adequately represent the interests of the members of the Class because the Plaintiffs have experienced the same harms as the members of the Class and have no conflicts with any other members of the Class. Furthermore, Plaintiffs have retained sophisticated and competent counsel who are experienced in prosecuting federal and state class actions, as well as other complex litigation.

**E. Commonality and Predominance**

~~88~~87. This action is appropriate as a class action under Rule 23(b)(3) of the Federal Rules of Civil Procedure because common questions of law predominate over individual questions including:

- a. Whether Defendants' copying and downloading of the Class's copyrighted works and using them to train the Llama 1 and Llama 2 language models constitute copyright infringement;

- b. Whether any statutes of limitation limits Plaintiffs' and the Class's potential for recovery;
- c. Whether Defendants' copying and downloading of the Class's copyrighted works was fair use;
- d. Whether Class members were harmed by Meta's copying and downloading of the Class's Works to train the Llama 1 and Llama 2 language models, and whether Class members are entitled to damages, including statutory damage and the amount of such damages.

~~89.88.~~ These and other questions of law and fact are common to the Class predominate over any questions affecting the members of the Class individually.

**F. Other Class Considerations**

~~90.89.~~ Defendants have acted on grounds generally applicable to the Class. This class action is superior to alternatives, if any, for the fair and efficient adjudication of this controversy. Prosecuting the claims pleaded herein as a class action will eliminate the possibility of repetitive litigation. There will be no material difficulty in the management of this action as a class action.

~~91.90.~~ The prosecution of separate actions by individual Class members would create the risk of inconsistent or varying adjudications, establishing incompatible standards of conduct for Defendants.

**VIII. DEMAND FOR JUDGMENT**

Wherefore, Plaintiffs request that the Court enter judgment on their behalf and on behalf of the Class defined herein, by ordering:

- a) This action may proceed as a class action, with Plaintiffs serving as Class Representatives, and with Plaintiffs' counsel as Class Counsel. Judgment in favor of Plaintiffs and the Class and against Defendant.

- b) A declaration that Meta has infringed Plaintiffs and the Class' exclusive copyrights in the Infringed Works under the Copyright Act.
- c) A declaration that such infringement is willful.
- d) An award of Plaintiffs' and the Class' actual damages and profits under 17 U.S.C. § 504(b) as shall be determined at trial, or at their election, an award of statutory damages in an amount to be determined at trial, as provided in 17 U.S.C. § 504(c), resulting from Meta's willful infringement of Plaintiffs' and the Class' exclusive copyrights in the Infringed Works.
- e) An order of costs and allowable attorneys' fees under 17 U.S.C. § 505.
- a) Pre- and post-judgment interest on the damages awarded to Plaintiffs and the Class, and that such interest be awarded at the highest legal rate from and after the date this class action complaint is first served on Defendant.
- b) Defendants are to be jointly and severally responsible financially for the costs and expenses of a Court approved notice program through post and media designed to give immediate notification to the Class.
- c) Further relief for Plaintiffs and the Class as may be just and proper.

#### **IX. JURY TRIAL DEMANDED**

Under Federal Rule of Civil Procedure 38(b), Plaintiffs demand a trial by jury of all the claims asserted in this Complaint so triable.

Dated: ~~December 11, 2023~~ August 29, 2024

By: /s/ Joseph R. Saveri  
Joseph R. Saveri

Joseph R. Saveri (State Bar No. 130064)  
Cadio Zirpoli (State Bar No. 179108)  
Christopher K.L. Young (State Bar No. 318371)  
Holden Benon (State Bar No. 325847)

Aaron Cera (State Bar No. 351163)

Margaux Poueymirou (State Bar No. 356000) ~~Kathleen J. McMahon (State Bar No. 340007)~~

**JOSEPH SAVERI LAW FIRM, LLP**

601 California Street, Suite 1050 ~~05~~

San Francisco, California 94108

Telephone: (415) 500-6800

Facsimile: (415) 395-9940

Email: jsaveri@saverilawfirm.com

czirpoli@saverilawfirm.com

cyoung@saverilawfirm.com

hbenon@saverilawfirm.com

acera@saverilawfirm.com

mpoueymirou@saverilawfirm.com ~~kmcme~~

mahon@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)

1920 Hillhurst Avenue, #406

Los Angeles, CA 90027

Telephone: (323) 968-2632

Facsimile: (415) 395-9940

Email: mb@buttericklaw.com

Bryan L. Clobes (*pro hac vice*)

**CAFFERTY CLOBES MERIWETHER  
& SPRENGEL LLP**

135 South LaSalle Street, Suite 3210

Chicago, IL 60603 ~~205 N. Monroe Street~~

Media, PA 19063

Telephone: (215) 864-2800

Email: bclobes@caffertyclobes.com

asweatman@caffertyclobes.com

Daniel J. Muller (State Bar No. 193396)  
**VENTURA HERSEY & MULLER, LLP**  
1506 Hamilton Avenue  
San Jose, California 95125  
Telephone: (408) 512-3022  
Facsimile: (408) 512-3023  
Email: [dmuller@venturahersey.com](mailto:dmuller@venturahersey.com)

Seth Haines\*  
Timothy Hutchinson\*  
Lisa Geary\*  
**RMP, LLP**  
5519 Hackett Street, Suite 300  
Springdale, Arkansas 72762  
Telephone: (479) 443-2705  
Email: [shaines@rmp.law](mailto:shaines@rmp.law)

David A. Straite (*pro hac vice*)  
**DICELLO LEVITT LLP**  
475 Lexington Avenue, Suite 1001  
New York, NY 10017  
Telephone: (646) 933-1000  
Email: [dstraite@dicellolevitt.com](mailto:dstraite@dicellolevitt.com)  
[thutchinson@rmp.law](mailto:thutchinson@rmp.law)  
[lgeary@rmp.law](mailto:lgeary@rmp.law)

Scott Poynter\*  
**POYNTER LAW GROUP**  
407 President Clinton Avenue, Suite 201  
Little Rock, Arkansas 72201  
Telephone: (501) 812-3943  
Email: [scott@poynterlawgroup.com](mailto:scott@poynterlawgroup.com)

Brian O'Mara  
**DiCELLO LEVITT LLP**  
4747 Executive Drive, Second Floor  
San Diego, California 92121  
Telephone: (619) 923-3939  
Email: [bomara@dicellolevitt.com](mailto:bomara@dicellolevitt.com)

David A. Straite (*pro hac vice*)  
**DiCELLO LEVITT LLP**  
475 Lexington Avenue, Suite 1001  
New York, NY 10017  
Telephone: (646) 933-1000  
Email: [dstraite@dicellolevitt.com](mailto:dstraite@dicellolevitt.com)

Adam J. Levitt

Amy E. Keller\*

Nada Djordjevic\*

James A. Ulwick\*

**DICELLO LEVITT LLP**

Ten North Dearborn Street, Sixth Floor

Chicago, Illinois 60602

Telephone: (312) 214-7900

Email: [alevitt@dicellolevitt.com](mailto:alevitt@dicellolevitt.com)

[akeller@dicellolevitt.com](mailto:akeller@dicellolevitt.com)

[ndjordjevic@dicellolevitt.com](mailto:ndjordjevic@dicellolevitt.com)

[julwick@dicellolevitt.com](mailto:julwick@dicellolevitt.com)

*Counsel for Individual and Representative  
Plaintiffs and the Proposed Class*